Indonesian Journal of Applied Statistics

E-ISSN 2621–086X

Vol. 6, No. 2, pp. **186-197**, **November 2023**
https://jurnal.uns.ac.id/ijas/article/view/80118
https://doi.org/10.13057/ijas.v6i2.80118

Pemodelan Data Time Series Menggunakan Pendekatan Regresi Polinomial Lokal Pada Data Harga Saham MDKA

Febrian Adri Nur Fauzi[1], Rukun Santoso[2*], Di Asih I Maruddani[3]
[123]Departemen Statistika, Universitas Diponegoro, Semarang, Indonesia

***Corresponding author**: rukunsantoso25@gmail.com

Abstract. *Investment is an important way to manage finances for profit. One of the most popular investments in Indonesia is buying and selling shares. In addition to getting profits, they also have risks. Therefore, analyzing stock prices before buying and selling is an important key in stock investing. Investors should buy stocks at a low price and sell them at a high price. One of the methods used is parametric regression analysis, but it has assumptions that must be met. A more flexible alternative is local polynomial regression without any particular assumptions. PT Merdeka Copper Gold Tbk with MDKA stock code is a company engaged in the mining and industrialization of gold, silver, and other associated minerals. The study of modeling the lowest daily price of MDKA shares using local polynomial regression showed excellent results. The high coefficient of determination exceeding 67% on the in-sample data indicates strong model performance, and the Mean Absolute Percentage Error (MAPE) value on the out-of-sample data is less than 10%, ensuring excellent model accuracy*.

Keywords: *local polynomial regression; MDKA shares; time series*

1. PENDAHULUAN

Investasi adalah salah satu cara untuk mengelola keuangan**.** Tujuan dari investasi adalah memperoleh keuntungan. Saham merupakan salah satu investasi yang popular di Indonesia. Namun, disamping memperoleh keuntungan, investasi saham juga memiliki risiko yang harus diwaspadai. Untuk memaksimalkan keuntungan dan meminimalkan risiko tersebut maka perlu dilakukan analisis harga saham sebelum membeli dan saat hendak menjual saham. Strategi ini melibatkan pembelian pada harga terendah dan penjualan pada harga tertinggi.

Analisis regresi adalah suatu metode untuk pemodelan data. Ada 2 pendekatan dalam pemodelan data yaitu pendekatan parametrik dan nonparametrik. Pada model regresi parametrik terdapat syarat dan asumsi-asumsi yang harus terpenuhi yaitu residualnya saling bebas, berdistribusi Gaussian (normal) dengan mean nol dan tidak bersifat heteroskedastis. Khususnya jika datanya berupa *time series*, pemodelan parametrik yang popular adalah *Autoregressive Integrated Moving Average* (ARIMA) Box Jenkins yang menyaratkan datanya harus stasioner dan asumsi residual seperti yang telah disebutkan. Jika asumsi-asumsi tersebut tidak dipenuhi, maka kesimpulan pada model parametrik akan menyebabkan tidak valid [1] . Oleh sebab itu, diperlukan pemodelan alternatif yang lebih fleksibel tanpa memerlukan adanya asumsi residual tersebut yaitu model nonparametrik dan salah satunya adalah model regresi polinomial lokal. Suparti dan Prahutama [2] dan Hendrian et al. [3] telah memodelkan data *time series* yaitu data

Submitted: 06 September 2023; ***Accepted:*** 28 Februari 2024

pemakaian beban listrik Kota Semarang dan harga emas dunia menggunakan regresi polinomial lokal. Penelitian tersebut tidak melibatkan sifat stasioneritas data dan tidak melakukan pengukuran terhadap kebaikan model, sehingga metode tersebut masih perlu dilakukan penelitian lebih lanjut.

Pemodelan data harga saham PT Merdeka Copper Gols Tbk. (MDKA) dipilih karena dari studi pendahuluan yang telah dilakukan menunjukkan pergerakan harga saham memiliki kecenderungan naik dalam 5 tahun terakhir, sehingga diharapkan akan memberikan keuntungan yang lebih tinggi dalam investasi. Analisis dilakukan pada harga terendah harian saham MDKA karena harga tersebut mencerminkan harga minimum yang berpotensi sebagai titik masuk untuk melakukan pembelian saham. Data tersebut tidak stasioner, namun setelah dilakukan penanganan stasioneritas, data memiliki pola regresi yang menggerombol pada beberapa titik. Analisis regresi polinomial lokal secara teori cocok untuk diterapkan pada data tersebut karena model regresi polinomial lokal mengadopsi perluasan Deret Taylor yang merepresentasikan suatu fungsi polinomial di sekitar titik pusat yang dinamakan titik lokal [4]. Tujuan dari penelitian ini adalah memodelkan data *time series* harga harian terendah saham MDKA menggunakan pendekatan regresi polinomial lokal dengan memperhatikan sifat stasioneritas data.

2. TINJAUAN PUSTAKA

Harga saham adalah harga jual beli yang ditetapkan di pasar saham untuk satu lembar saham dari sebuah perusahaan yang telah terdaftar pada bursa saham. Harga saham merupakan data *time series*. Pemodelan data *time series* memerlukan syarat stasioner pada datanya baik dalam mean maupun varian yaitu mean dan variannya konstan sepanjang waktu [5]. Proses stasioner dapat diasumsikan seperti pada Persamaan (1) dan (2).

$$E(Z_t) = \mu \tag{1}$$

$$Var(Z_t) = E(Z_t - \mu)^2 = \gamma_0 \tag{2}$$

Untuk menguji data stasioner dalam *mean* dapat dilakukan dengan uji akar unit *Dickey-Fuller* (DF). Uji akar unit menggunakan model regresi yang dituliskan pada Persamaan (3) [6].

$$\Delta Z_t = (\rho - 1)Z_{t-1} + \varepsilon_t$$

$$\Delta Z_t = \delta Z_{t-1} + \varepsilon_t \tag{3}$$

dengan $\delta = \rho - 1$ dan $\Delta Z_t = Z_t - Z_{t-1}$.

Z_t dikatakan tidak stasioner (memiliki akar unit) apabila $\delta = 0$ yang berarti $\rho = 1$

Hipotesis untuk uji akar unit *Dickey-Fuller* (DF) yaitu:

H_0 : $\delta = 0$ (Data tidak stasioner dalam *mean*)

H_1 : $\delta < 0$ (Data stasioner dalam *mean*)

Statistik ujinya dinyatakan oleh Persamaan (4).

$$\tau_{hitung} = \frac{\hat{\delta} - \delta}{SE(\hat{\delta})} \tag{4}$$

dengan $SE(\hat{\delta}) = \sqrt{\frac{\sigma^2}{\sum_{t=2}^{n} (Z_{t-1} - \underline{Z}_{t-1})}}$, $\hat{\delta}$: estimasi dari δ, dan $SE(\hat{\delta})$: *standard error* dari δ.

Kriteria uji pada uji DF adalah H_0 ditolak jika nilai $\tau_{hitung} < \tau_{(\alpha;n-1)}$ atau $p - value < \alpha$ Jika data *time series* tidak stasioner dalam *mean*, untuk menstsionerkan data tersebut dapat dilakukan diferensiasi. Proses diferensiasi orde ke-d dapat dinyatakan oleh Persamaan (5).

$$\Delta^d Z_t = (1 - B)^d Z_t \tag{5}$$

dengan $\Delta^d z_t$ adalah hasil diferensiasi pengamatan ke–t orde ke-d, d adalah orde diferensiasi, dan B adalah operator *backshift.*

Pengujian stasioneritas dalam varian dapat melihat nilai λ pada transformasi Box-Cox. Nilai λ yang tidak sama dengan satu menandakan bahwa data tidak stasioner dan harus dilakukan transformasi Box-Cox yang dinyatakan oleh Persamaan (6) [5].

$$Z_t^* = \{\frac{Z_t^\lambda - 1}{\lambda}, \quad \lambda \neq 0 \; ln \; Z_t, \quad \lambda = 0 \tag{6}$$

Regresi polinomial lokal merupakan implementasi deret Taylor di sekitar x_0. Representasi deret Taylor dari g(X) di sekitar x_0 dituliskan pada Persamaan (7).

$$g(X) = g(x_0) + g'(x_0)(X - x_0) + \frac{g''(x_0)}{2!}(X - x_0)^2 + \cdots + \frac{g^{(m)}(x_0)}{m!}(X - x_0)^m + o((X - x_0)^{m+1}) \tag{7}$$

Model regresi polinomial lokal derajat m atau order m+1 disajikan pada Persamaan (8) [4].

$$Y_t = g(X_t) + \varepsilon_t \tag{8}$$

dengan $g(X_t) = \sum_{j=0}^{m} \beta_j (X_t - x_0)^j$, $\beta_j = \frac{g^{(j)}(x_o)}{j!}$, j = 0,1, …, m dan t =1,2, …, n.

Model pada Persamaan (8) dapat dinyatakan dalam bentuk matriks seperti yang ditunjukkan dalam Persamaan (9).

$$Y = X\beta + \varepsilon \tag{9}$$

dengan:

$$\boldsymbol{X} = \begin{pmatrix} 1 & (X_1 - x_0) & (X_1 - x_0)^2 & \ldots & (X_1 - x_0)^m \\ 1 & (X_2 - x_0) & (X_2 - x_0)^2 & \ldots & (X_2 - x_0)^m \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & (X_n - x_0) & (X_n - x_0)^2 & \ldots & (X_n - x_0)^m \end{pmatrix}$$

$$\boldsymbol{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} ; \; \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_m \end{pmatrix} \text{ dan } \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}.$$

Parameter β pada model Persamaan (9) dapat diestimasi dengan metode *Weighted Least Square* (WLS) menggunakan fungsi bobot kernel dan salah satunya adalah kernel Gaussian. Prinsip metode metode WLS adalah meminimumkan fungsi *Residual Sum of Square* (RSS) terboboti. Fungsi RSS ditulis pada Persamaan (10).

$$RSS = \sum_{t=1}^{n} \left\{ Y_t - \sum_{j=0}^{m} \beta_j (X_t - x_0)^j \right\}^2 K_h(X_t - x_0) \tag{10}$$

dengan h adalah lebar jendela (*bandwidth*) yang bernilai positip dan mengontrol titik lokal x_0. Bentuk matriks Persamaan (10) disajikan pada Persamaan (11)

$$RSS = (\boldsymbol{Y} - \boldsymbol{X\beta})^{\mathrm{T}} \boldsymbol{W} (\boldsymbol{Y} - \boldsymbol{X\beta}) \tag{11}$$

dengan T menyatakan *transpose* matriks dan W merupakan matriks bobot berbentuk diagonal ukuran n × n:

$$\boldsymbol{W} = \begin{pmatrix} K_h(X_1 - x_0) & 0 & \ldots & 0 \\ 0 & K_h(X_2 - x_0) & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & K_h(X_n - x_0) \end{pmatrix}$$

$$\begin{pmatrix} \frac{1}{h}K\left(\frac{X_1 - x_0}{h}\right) & 0 & \cdots & 0 \\ 0 & \frac{1}{h}K\left(\frac{X_2 - x_0}{h}\right) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{h}K\left(\frac{X_n - x_0}{h}\right) \end{pmatrix}$$

Solusi dari Persamaan (11) dapat dicari seperti pada Persamaan (12).

$$\hat{\beta} = (X^T W X)^{-1} X^T W Y \tag{12}$$

Estimasi model regresi polinomial lokal Persamaan (8) atau Persamaan (9) dituliskan dalam Persamaan (13).

$$\hat{Y}_t = \hat{g}(X_t) = \sum_{j=0}^{m} \hat{\beta}_j (X_t - x_0)^j \quad \text{atau}$$

$$\hat{\boldsymbol{Y}} = \hat{g}(\boldsymbol{X}) = \boldsymbol{X}\hat{\boldsymbol{\beta}} = \boldsymbol{X}\left(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{Y} = \boldsymbol{H}\boldsymbol{Y} \tag{13}$$

dengan $\boldsymbol{H} = \boldsymbol{X}\left(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{W}$ disebut matriks Hat.

Pasangan $\{(X_t, Y_t), t = 1, 2, \dots, n\}$ dalam pemodelan regresi adalah saling bebas. Dalam praktiknya banyak ditemukan asumsi kebebasan data tidak terpenuhi. Contohnya, ketika mengamati data yang disusun dalam urutan waktu dari suatu objek penelitian, respons objek saat ini dapat dipengaruhi oleh respons sebelumnya. Oleh karena itu, diperlukan pengembangan suatu model yang mempertimbangkan ketidakbebasan data tersebut. Hardle (1990) [7] memberikan konsep matematika yang menjadi dasar dalam pengembangan model, yaitu:

1. Model (S): Suatu barisan stasioner $\{(X_t, Y_t), t = 1, 2, \dots, n\}$ (antar pengamatan diijinkan tak bebas stokastik) telah terobservasi dan akan diestimasi $\hat{Y}_t$ menggunakan $g(x) = E(Y|X = x)$.
2. Model (Ts): Suatu data *time series* $\{Z_t, t \geq 1\}$ telah terobservasi dan akan diprediksi Z_{n+1} dengan $g(x) = E(Z_{n+1}|Z_n = x)$.

Permasalahan model (Ts) dapat ditetapkan ke dalam model (S) dengan mengubah *time series* $\{Z_t, t \geq 1\}$, nilai lag Z_{t-1} sebagai X_t dan Z_t sebagai Y_t. Selanjutnya untuk memprediksi Z_{n+1} dari $\{Z_t\}$ dapat dipandang sebagai masalah pemodelan regresi untuk $\{(X_t, Y_t)\} = \{(Z_{t-1}, Z_t), t = 1, 2, \dots, n\}$. Jadi untuk memprediksi $\{Z_t\}$ ekivalen dengan melakukan estimasi $g(x) = E(Y|X = x)$ untuk data *time series* dua dimensi $\{(X_t, Y_t), t = 1, 2, \dots, n\}$.

Baik atau buruknya model polinomial lokal dipengaruhi oleh fungsi bobot kernel, derajat polinomial, titik lokal, dan nilai *bandwidth*-nya. Nilai *bandwidth* yang besar akan menghasilkan estimasi model sangat mulus (*oversmoothing*), dengan bias pemodelan yang besar dan varian yang kecil. Di sisi lain, jika nilai *bandwidth* terlalu kecil, estimasi model akan menjadi sangat berfluktuasi (*undersmoothing*), dengan bias pemodelan yang kecil dan varian yang besar. Nilai *bandwidth* akan mengontrol titik lokal xo. Selain *bandwidth*, pemilihan derajat polinomial lokal juga memiliki pengaruh penting. Derajat yang lebih tinggi dapat mengurangi bias pemodelan tetapi juga meningkatkan varian. Maka dari itu, penting untuk menemukan nilai *bandwidth* dan derajat polinomial yang optimal yang dapat mencapai keseimbangan yang tepat antara bias dan varian, sehingga hasil estimasi dapat menjadi akurat. Dari parameter-parameter tersebut, pemilihan *bandwidth* adalah parameter yang paling dominan dalam menentukan model optimal [1, 2]. Metode GCV (*Generalized Cross Validation*) digunakan untuk optimasi agar mendapatkan h optimal. Metode GCV dirumuskan dalam Persamaan (14).

$$\mathrm{GCV}(h)=\frac{n^{-1}\sum_{t=1}^{n}[\mathrm{Y_t}-\hat{\mathrm{g}}(\mathrm{X_t})]^2}{\left[1-\frac{\sum_{t=1}^{n}[\mathrm{H}]tt}{n}\right]^2} \tag{14}$$

dengan Y_t : data pengamatan sebenarnya ke-t, $\hat{g}(X_t)$: hasil estimasi model polinomial lokal pada pengamatan X_t, H adalah matriks Hat, dan n adalah banyak pengamatan.

Koefisien determinasi digunakan untuk mengukur kebaikan model. Koefisien determinasi adalah sebuah metrik yang mengukur sejauh mana peubah bebas dapat menjelaskan variasi dalam peubah tak bebasnya. Persamaan (15) menggambarkan rumus koefisien determinasi [8].

$$\mathrm{R}^2=1-\frac{\sum_{\mathrm{t=1}}^{\mathrm{n}}\left(\mathrm{Z_t}-\hat{\mathrm{Z}}_{\mathrm{t}}\right)^2}{\sum_{\mathrm{t=1}}^{\mathrm{n}}(\mathrm{Z_t}-\bar{\mathrm{Z}})^2} \tag{15}$$

Nilai dari koefisien determinasi yang lebih dari 67% mengindikasikan model yang tergolong model yang kuat [9].

Ukuran kinerja model merupakan suatu ukuran untuk mengukur seberapa baik model jika diterapkan pada data *out sample*. Ukuran kinerja model dapat dideteksi dengan nilai MAPE (*Mean Absolute percentage error*). Nilai MAPE yang kurang dari 10 % mengindikasikan peramalan sangat akurat [10]. Nilai MAPE dirumuskan pada Persamaan (16)

$$\mathrm{MAPE}=\left(\frac{1}{\mathrm{n}}\sum_{\mathrm{t=1}}^{\mathrm{n}}\left|\frac{\mathrm{Z_t}-\hat{\mathrm{Z}}_{\mathrm{t}}}{\mathrm{Z_t}}\right|\right)\times 100\% \tag{16}$$

dengan Z_t adalah data pengamatan sebenarnya pada periode ke-t, $\hat{Z}_t$ adalah data hasil prediksi periode ke-t, dan n adalah banyak pengamatan

3. METODE PENELITIAN

Data penelitian ini adalah data sekunder dalam bentuk *time series* yang diambil dari situs web Yahoo Finance. Variabel dalam penelitian adalah harga harian terendah MDKA, dinotasikan sebagai Z_t. Data diambil mulai dari tanggal 4 Januari 2021 hingga 22 September 2022 yaitu sebanyak 422 data dengan perbandingan data *in sample* dan *out sample* adalah sebesar 90% untuk data *in sample* dan 10% untuk data *out sample*. *Software* R-Studio digunakan untuk melakukan pengolahan data pada penelitian ini.

Tahapan untuk melakukan pemodelan regresi polinomial lokal data *time series* yang pertama adalah menguji kestasioneritasan data. Uji kestasioneran data ada 2 tahap yaitu uji stasioneritas dalam varian dan mean. Pertama-tama dilakukan uji stasioneritas data dalam varian. Jika data belum stationer dalam varian dilakukan transformasi Box-Cox menjadi Z_t^*. Uji berikutnya adalah uji stasioneritas dalam mean. Jika data belum stasioner dalam mean dilakukan deferencing menjadi ΔZ_t^*. Langkah selanjutnya yaitu mengubah data *time series* yang telah stasioner ΔZ_t^* menjadi peubah tak bebas (Y) dan nilai lag ΔZ_{t-1}^* sebagai peubah bebas (X). Langkah selanjutnya membagi data menjadi 2 bagian yaitu data *in sample* dengan proporsi 90% dan data *out sample* dengan proporsi 10%. Data *in sampl*e digunakan untuk pemodelan regresi polinomial lokal dengan menentukan fungsi kernel, derajat polinomial, titik lokal, dan nilai *bandwidth*. Langkah selanjutnya yaitu melakukan estimasi parameter model dan menghitung nilai GCV untuk mencari nilai *bandwidth* optimal. Langkah selanjutnya yaitu menentukan model terbaik yang kemudian dihitung nilai koefisien determinasinya. Langkah selanjutnya yaitu melakukan prediksi data *out sample* yang kemudian dihitung nilai MAPE-nya. Langkah terakhir membuat kesimpulan.

4. HASIL DAN PEMBAHASAN

Data penelitian ini berupa data harga harian terendah saham MDKA mulai tanggal 4 Januari 2021 hingga 22 September 2022 yang disajikan pada Gambar 1.

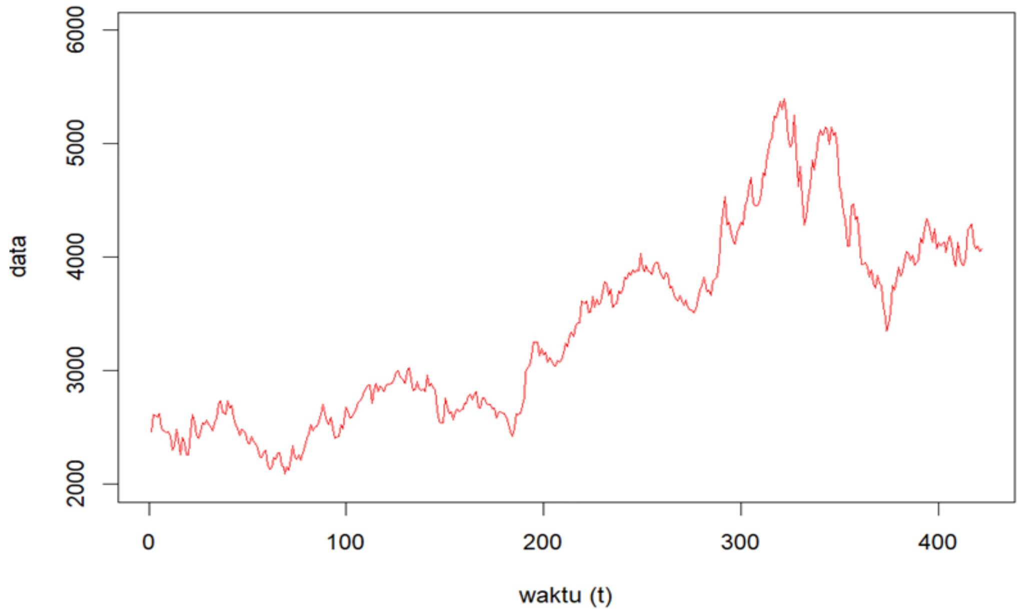
Pengujian stasioneritas dalam varian menunjukan nilai lamda pada transformasi Box-Cox sebesar -0,02 sehingga data harga saham terendah MDKA tidak stasioner dalam varian. Setelah dilakukannya transformasi diperoleh nilai lamda sebesar 1 sehingga data harga saham terendah MDKA setelah transformasi sudah stasioner dalam varian.

Langkah berikutnya adalah pengujian stasioneritas dalam *mean* menggunakan uji akar unit *Dickey Fuller* sebagai berikut:

Hipotesis:

H_0 : δ = 0 (Terdapat akar unit sehingga data *time series* tidak stasioner)
H_1: δ < 0 (Tidak terdapat akar unit sehingga data *time series* stasioner)

Taraf signifikansi: $\alpha = 0{,}05$
Statistik Uji: Hasil Uji akar unit *Dickey Fuller* disajikan pada Tabel 1.



Gambar 1. Plot Data *Time Series* Harga Saham Harian Terendah

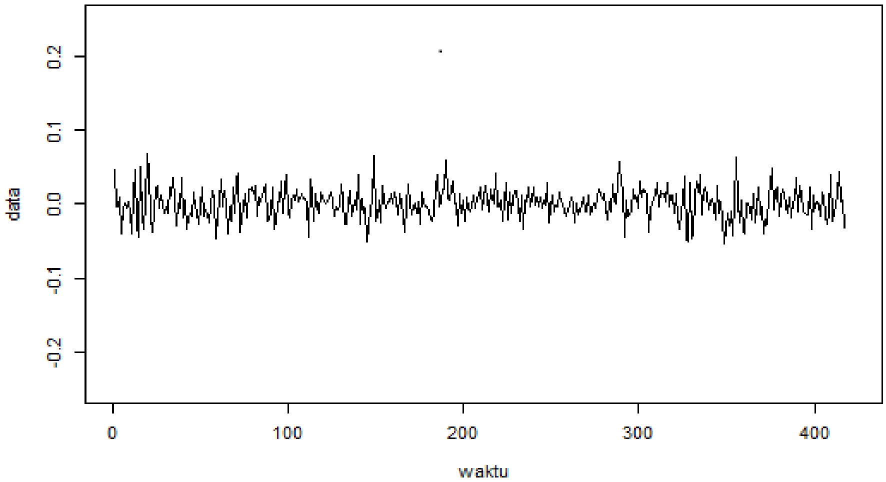
Tabel 1. Tabel Dickey Fuller Test Harga Saham Terendah MDKA

Data	τ	*P-value*	Keputusan
Sebelum *differencing*	$-2{,}2497$	0,472	H_0 diterima

Sesudah *differencing*	$-19{,}462$	$0{,}01$	H_0 ditolak

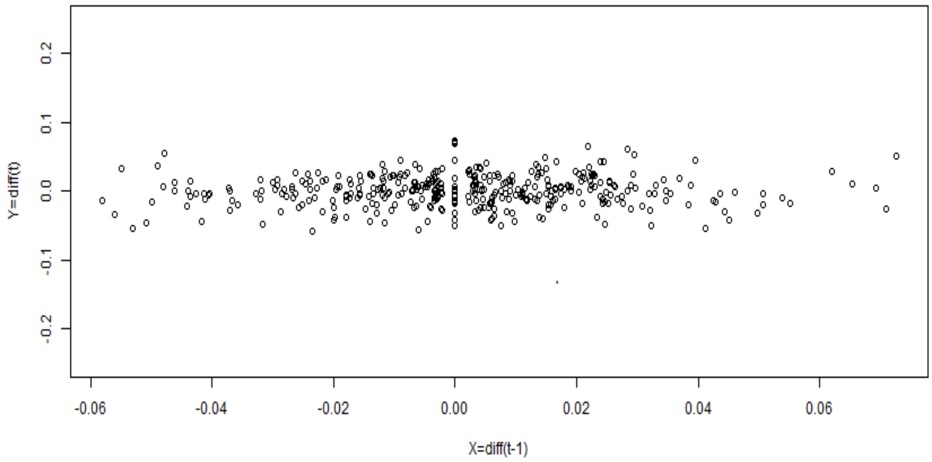
Kriteria Uji: Tolak H_0 jika $\tau <$ nilai kritis Dickey Fuller $(\alpha, n-1)$ atau *P-value* $< \alpha$ dengan $\alpha = 5\%$ dan dengan $n = 422$.

Kesimpulan: H_0 ditolak saat data di-*differencing* maka data stasioner dalam *mean* setelah di-*differencing* satu kali. Plot data yang telah stasioner disajikan pada Gambar 2.



Gambar 2. Plot Data Stasioner Harga Saham Harian Terendah

Data yang telah stasioner selanjutnya dimodifikasi menjadi data (X,Y). Data modifikasi dibagi menjadi 2 bagian yaitu data *in sample* dan *out sample* dengan perbandingan 90%:10%. Data *in sample* disajikan pada Gambar 3.



Gambar 3. Scatter Plot Data Modifikasi *In Sample* Stasioner Harga Saham Harian Terendah

Penentuan model dilakukan dengan menentukan *trial-error* dari derajat polinomial 0, 1, 2, dan 3, nilai titik lokal (x_0) yang berada pada rentang data X yaitu antara data minimum dan data maksimum, dan *bandwidth* dicobakan mulai dari 0,1 hingga 0,3 dengan penambahan sebesar 0,01 dan pencarian interval sebelumnya diubah-ubah untuk mendapatkan nilai GCV yang paling minimum sehingga diperoleh perhitungan nilai GCV terkecil seperti pada Tabel 2.

Tabel 2. Pemilihan *Bandwidth* Optimal

Fungsi Kernel	Derajat	*Bandwidth* Optimal	Titik Lokal	GCV Minimum
Gaussian	0	0,13	-0,018	0,0005156092
Gaussian	1	0,32	0,002	0,0005163823
Gaussian	2	0,23	0,052	0,0005153468
Gaussian	**3**	**0,28**	**0,062**	**0,0005153293**

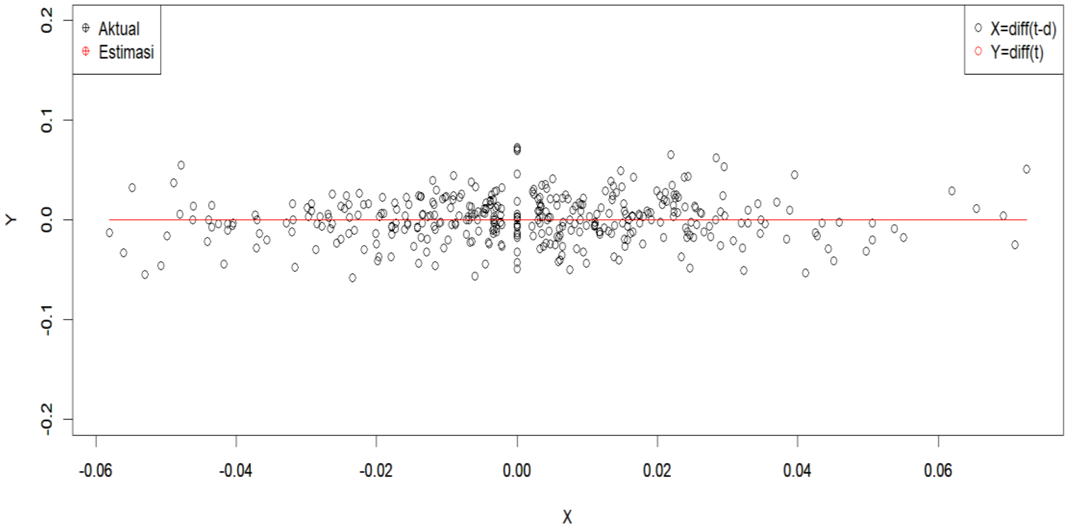
Berdasarkan Tabel 1 diperoleh model polinomial lokal terbaik dengan derajat 3, titik lokal $x_{10} = 0{,}062$, dan nilai *bandwidth* $h = 0{,}28$ dengan nilai GCV sebesar 0,0005153293. Estimasi model polinomial lokal terbaik disajikan pada Tabel 2.

Tabel 2. Estimasi Paremeter Model Terbaik

Parmeter	Hasil Estimasi
$\hat{\beta}_0$	-0,002798632
$\hat{\beta}_1$	-0,2242331
$\hat{\beta}_2$	-2,33346
$\hat{\beta}_3$	0,3996899

Berdasarkan Tabel 2, model regresi polinomial lokal yang terbentuk dituliskan dalam Persamaan (17). Plot data X dan Y aktual serta nilai estimasinya disajikan pada Gambar 4.

$$\hat{Y}_t = -0{,}002798632 - 0{,}2242331(X_t - 0{,}062) - 2{,}33346(X_t - 0{,}062)^2 + 0{,}3996899(X_t - 0{,}062)^3 - (4{,}62296e - 06)(X_{1t} - 0{,}062)^3 \quad (17)$$



Gambar 4. Plot data X dan Y Aktual dan Nilai Estimasinya

Model polinomial lokal yang diperoleh pada Gambar 4 adalah model dari data yang telah distasionerkan, terlihat bahwa plot nilai estimasi mengikuti nilai data aktual. Agar model dapat digunakan untuk melakukan estimasi data aslinya, maka estimasi $\hat{Y}_t$ perlu dikembalikan ke bentuk aslinya yaitu melakukan invers dari *diferensiasi dan* transformasinya. Model regresi polinomial lokal (17) dikembalikan ke data aslinya dengan mengganti:

$$\mathrm{Y_t} = \Delta \mathrm{Z^*}_\mathrm{t} = \mathrm{Z^*}_\mathrm{t} - \mathrm{Z^*}_{\mathrm{t-1}} \text{ dengan } \mathrm{Z^*}_\mathrm{t} = \frac{\mathrm{Z_t}^{\lambda}-1}{\lambda} \tag{18}$$

$$\mathrm{X_t} = \Delta \mathrm{Z^*}_{\mathrm{t-1}} = \mathrm{Z^*}_{\mathrm{t-1}} - \mathrm{Z^*}_{\mathrm{t-2}} \tag{19}$$

Dari persamaan (18) diperoleh Persamaan (20)

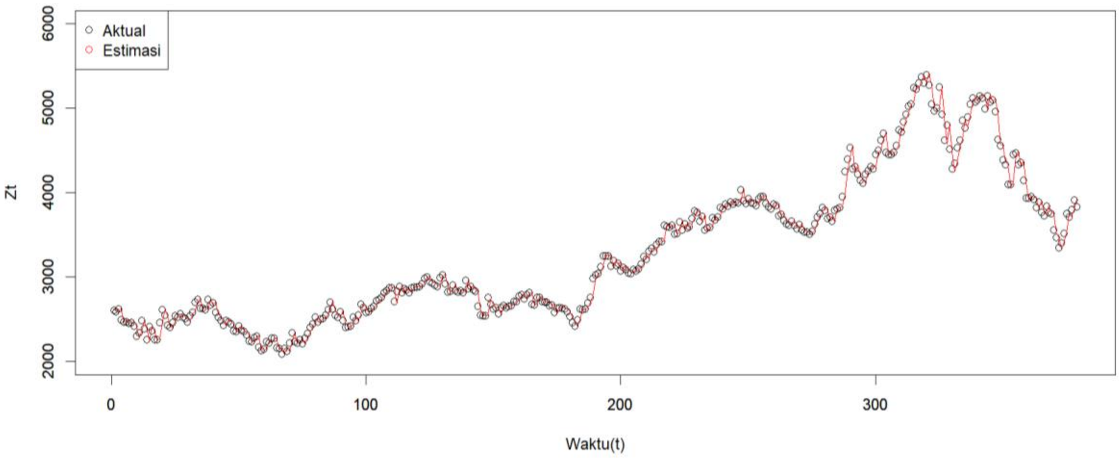
$$\hat{\mathrm{Y}}_\mathrm{t} = \widehat{\mathrm{Z^*}}_\mathrm{t} - \mathrm{Z^*}_{\mathrm{t-1}} \text{ sehingga } \widehat{\mathrm{Z^*}}_\mathrm{t} = \hat{\mathrm{Y}}_\mathrm{t} + \mathrm{Z^*}_{\mathrm{t-1}} \tag{20}$$

Dari Persamaan (18), $\mathrm{Z^*}_\mathrm{t} = \frac{\mathrm{Z_t}^{\lambda}-1}{\lambda}$ diperoleh $\widehat{\mathrm{Z^*}}_\mathrm{t} = \frac{\widehat{\mathrm{Z_t}}^{\lambda}-1}{\lambda}$ dan $\widehat{\mathrm{Z}}_\mathrm{t} = \left(\lambda\widehat{\mathrm{Z^*}}_\mathrm{t} + 1\right)^{1/\lambda}$.

Dari Persamaan (19) dan Persamaan (20), diperoleh pemodelan harga terendah harian saham MDKA pada Persamaan (21).

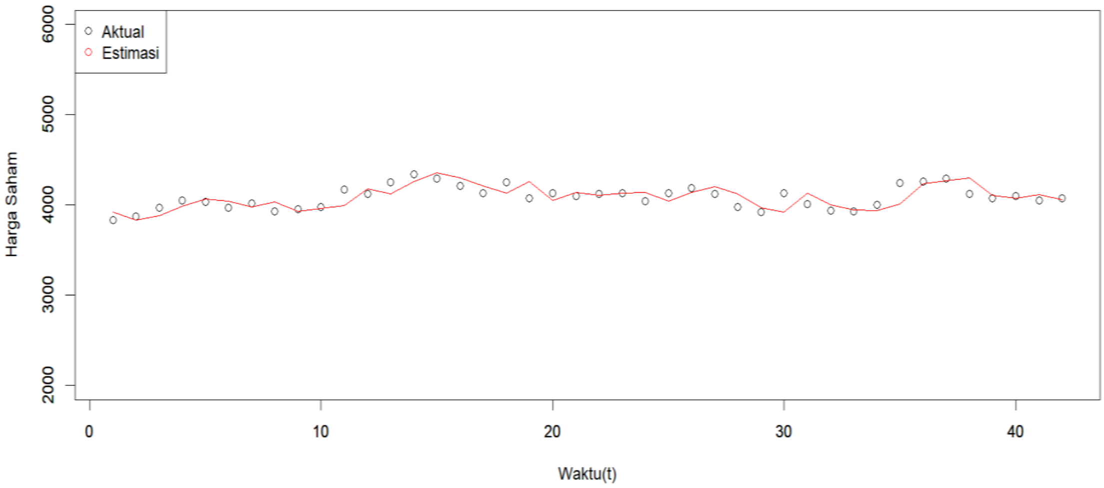
$$\begin{aligned}\hat{\mathrm{Z}}^*{}_\mathrm{t} &= -0{,}002798632 + \mathrm{Z^*}_{t-1} - 0{,}2242331(\mathrm{Z^*}_{\mathrm{t-1}} - \mathrm{Z^*}_{\mathrm{t-2}} - 0{,}062) \\ &\quad -2{,}33346(\mathrm{Z^*}_{\mathrm{t-1}} - \mathrm{Z^*}_{\mathrm{t-2}} - 0{,}062)^2 + 0{,}3996899(\mathrm{Z^*}_{\mathrm{t-1}} - \mathrm{Z^*}_{\mathrm{t-2}} - 0{,}062)^3\end{aligned} \tag{21}$$

Plot data sebenarnya dan estimasi yang telah dikembalikan ke bentuk data *time series* aslinya disajikan pada Gambar 5.



Gambar 5. Plot Data *In Sample* Harga Saham Harian Terendah Aktual dan Estimasinya

Gambar 5 menunjukkan bahwa pendekatan estimasi digambarkan dengan baik melalui visualisasi plot data asli. Permodelan harga saham terendah saham ini ditunjukkan dengan baik dalam penggunaan pemodelan polinomial lokal. Berdasarkan hasil perhitungan juga diperoleh informasi bahwa nilai R^2 sebesar 0,9885946 artinya bahwa peubah bebas memberi kontribusi pada variabilitas peubah tak bebas sebesar 98,86% dan ini mengindikasikan bahwa modelnya termasuk kriteria model kuat karena lebih besar dari 67% sehingga model memiliki kemampuan yang baik secara *in sample*. Namun kemampuan yang baik secara *in sample* belum tentu model juga memiliki kinerja yang baik untuk melakukan estimasi pada data *out sample*, sehingga perlu dihitung nilai MAPE *out sample* untuk mengetahui ukuran kinerja model sangat baik, baik, cukup, atau buruk dengan penghitungan dari data sebenarnya *out sample* dan data estimasi *out sample.* Grafik perbandingan data *out sample* sebenarnya harga saham terendah dan estimasinya disajikan pada Gambar 6.



Gambar 6. Grafik Data *Out Sample* Aktual dan Estimasinya

Gambar 6 menggambarkan bahwa plot data asli *out sample* sangat dekat hasil estimasinya. Berdasarkan perhitungan yang telah dilakukan, model memiliki MAPE *Out Sample* sebesar

2,039%, artinya bahwa kesalahan model untuk meramalkan data *out sample* adalah 2,039%. Kemampuan estimasi data *out sample* sangat baik dapat ditunjukkan melalui nilai MAPE < 10%

5. KESIMPULAN

Hasil penelitian dapat disimpulkan bahwa pemodelan harga harian terendah saham MDKA menggunakan polinomial lokal dengan pembobot fungsi kernel Gaussian diperoleh model terbaik mempunyai derajat 3 dengan titik lokal sebesar 0,062 yang dinyatakan pada Persamaan

$$\widehat{Z_t} = \left(\lambda\widehat{Z^*}_t + 1\right)^{1/\lambda}$$

dengan:

$$\hat{Z}^*_t = -0{,}002798632 + Z^*_{t-1} - 0{,}2242331(Z^*_{t-1} - Z^*_{t-2} - 0{,}062)$$
$$-2{,}33346(Z^*_{t-1} - Z^*_{t-2} - 0{,}062)^2 + 0{,}3996899(Z^*_{t-1} - Z^*_{t-2} - 0{,}062)^3$$

$$Z^*_t = \frac{Z_t^{\lambda} - 1}{\lambda} \text{ dan } \lambda = -0{,}02.$$

Model terbaik mempunyai koefisisen determinasi sebesar 0,9824 artinya bahwa peubah bebas memberi kontribusi pada variabilitas peubah tak bebas sebesar 98,24% dan model termasuk kriteria kuat karena koefisien determinasi lebih dari 67%. Model terbaik juga memiliki kinerja yang sangat bagus karena MAPE out sample sebesar 2,039% kurang dari 10%. Penelitian ini menghasilkan pemodelan harga harian terendah saham MDKA sebagai referensi dalam berinvestasi, terutama menjadi strategi untuk memilih waktu yang tepat dalam membeli saham dengan harga terendah. Dengan memasukkan data harga harian terendah 2 periode terakhir dapat dilakukan prediksi harga saham terendah periode-periode berikutnya.

DAFTAR PUSTAKA

[1] Suparti, R. Santoso, A. Prahutama, A. R. Devi, Regresi Nonparametrik. Ponorogo Indonesia: Wade Group, 2018.

[2] Suparti dan A. Prahutama, “Pemodelan Regresi Nonparametrik Menggunakan Pendekatan Polinomial Lokal Pada Beban Listrik Di Kota Semarang,” Media Statistika, vol. 9, no.2, pp. 85–93, 2016.

[3] J. Hendrian, Suparti, A. Prahutama, “Pemodelan Harga Emas Dunia Menggunakan Metode Nonparametrik Polinomial Lokal Dilengkapi GUI R,” Jurnal Gaussian, vol. 10, no. 4, pp. 604-616, 2021.

[4] K. Takezawa, Introduction to Nonparametric Regression. John Wiley & Son. Inc., 2005.

[5] G. E. P. Box, G. M. Jenkins, G. C. Reinsel and G. M. Ljung, Time Series Analysis Forecasting and Control Fifth Edition. John Wiley & Sons.Inc., 2016.

[6] D. A. Dickey dan W. A. Fuller, “ Distribution of Estimators for Autoregressive Time Series with a Unit Root,” Journal America Statistical Association, vol. 74, pp. 427-443, 1979.

[7] W. Hardle, Applied nonparametric regression. New York: Cambridge University Press., 1990.

[8] D. Chicco, M. J. Warrens, dan G. Jurman, “ The Coefficient of Determination R-squared is More Informative than SMAPE, MAE, MAPE, MSE, and RMSE in Regression Analysis Evaluation,” PeerJ Compt. Sci., vol. 7, no. 3, pp. 1-24, 2021.

[9] W. W. Chin, The Partial Least Square Aproach to Structural Equation Modeling. London: Lawrence Erlbaum Associates, 1998.

[10] P. C. Chang, Y. W. Wang, dan C. H. Liu, "The Development of a Weighted Evolving Fuzzy Neural Network for PCB Sales Forecasting," Expert Systems with Application, vol .32, pp. 88-89, 2007.